Trust and Detection:
An Experimental Investigation of Motivational Crowding Out

Matteo Ploner

CEEL Working Paper N. **02/2005**

CEEL - Computable and Experimental Economics Laboratory

Via Inama, 5 38100 Trento -Italy-

tel. +39.0461.882246 - fax. +39.0461.882222

http://www-ceel.economia.unitn.it/

Trust and Detection:
An Experimental Investigation of Motivational Crowding Out

Matteo Ploner*

March 2006

1 Introduction

Standard agency relationships are structured around divergent interests on the side of the Principal and of the Agent. When a Principal does not trust an Agent there are strong incentives for the former to impose bounds on the Agent's actions or to discover the actual type of those with whom interacting. Prendergast (1999) notices, with reference to the standard approach of economics to agency problems, that "*incentives are provided to workers through the compensation practices of firms, encompassing monitoring, evaluation, and contracting and firms use many different mechanisms to align interests*". However, recent experimental contributions have highlighted the importance of incentive structures which are not considered by the standard framework of selfish rational approach (Fehr *et al.*, 1998). Aim of the present contribution is to investigate, through an experimental analysis, the

*LEM, Sant'Anna School of Advanced Studies, Pisa (Italy)
e-mail address: ploner@sssup.it
homepage: https://mail.sssup.it/~ploner

relationship between trust, intrinsic motivations and autonomy supportive behavior in a hierarchical two-party relationship. The main hypothesis leading experimental work is that tasks which are freely managed by Agents (i.e., not under the control of the Principal), are more likely to foster intrinsic motivations and that environments where intrinsic motivations prevail will favor cooperative behavior.

The paper focuses on choices signaling trust and trustworthiness. Trust, as a reflection of social capital, is a constitutive element of economic transactions (Arrow, 1974). A considerable amount of economic contributions has focused on the impact of social capital on economic growth (Porta *et al.*, 1997). Many different conceptualizations of trust have been proposed (Hardin, 2001) but the dominant definition in the economic literature is still the one based on rational maximization (i.e., trust as calculativeness). As Williamson (1993) notices in a well-known work, calculative trust is not formally distinguishable from the standard economic paradigm of utility maximization under risk. Unfortunately, a trust process based on strict risk-benefit analysis has to be considered, quoting Williamson, as "a contradiction in terms"[p. 463]. Other works have rejected the strict consequentialist characterization of trust provided by mainstream economics and have preferred to switch to content-specific trust lead by social heuristics. According to this kind of conceptualization, trust is an "appropriate" (March, 1994) behavior matching characteristics of the subject and the environment. Some of these contributions (see among others, Messik and Kramer, 2003;

Yamaghishi, 2001) have evidenced the role of the reference group and identification with this in defining trust and trustworthy behavior. In recent times, experimental evidence showing that a standard definition of calculativeness trust cannot explain the behavior of subjects in very simple games (Berg *et al.*, 1995; Cox, 2004) has been collected. Cultural and social factors have been considered in explaining trust and reciprocity behavior in experimental settings (Glaeser *et al.*, 2000; Buchan and Croson, 2004). Other experimental contributions have found that trust is more influenced by innate characteristics of the subjects (Gächter *et al.*, 2004) and that artificial manipulation of the relationship between the subjects does not provide significant results in rewarded laboratory experiments (Güth *et al.*, 2004). For what attains trust in laboratory settings which try to replicate workplace environments, it is interesting to refer to the work of Fehr *et al.* (1998). The experiments conducted show that higher offers in terms of retribution are honored by workers even if the contract to which they adhere is incomplete and the Principal cannot enforce the agreement. The present study analyzes the interaction between intrinsic motivations and social preferences in the form of trust and reciprocity. Intrinsic motivations have been defined as those motivations associated with doing something "because it is inherently interesting or enjoyable"(Ryan and Deci, 2000). As shown by previous contributions autonomy in decision making is the key element for intrinsic motivations to prevail. In this perspective a control strategy introduced by the Principal to bound the actions of the Agent is likely to be perceived as a signal of distrust (on this aspect see, among others, Deci *et al.*, 1989). For what attains real-life observations, perceived distrust in the workplace has been shown to destroy cooperation and intrinsic motivations of the employees (Kramer, 1999).

In previous studies a lot of attention has been paid to individual and environmental characteristics as determinants of trust and reciprocity. At the same time, the propensity of economics to consider preferences as given primitives has led to the neglection of processes characterizing preference formation. The present work aims at considering, with a certain degree of novelty, the interaction between trust preferences and intrinsic motivations of the subjects. Particular attention will be reserved for the impact of a control strategy on reciprocity and intrinsic motivations when confronted with an alternative autonomy-supportive strategy. The experimental design which will be presented in section 2 has been conceived to account, within a simple and manageable structure, for choices of the Agent characterized alternatively by intrinsic or extrinsic motivations in association with autonomy-bounding or autonomy-supportive actions performed by the Principal.

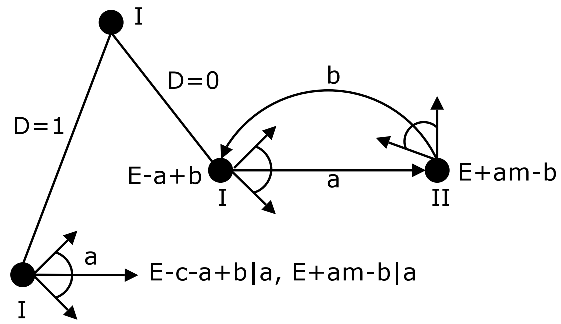
2 Experimental Design

The main consideration when designing the experiment was the possibility of testing alternative behavior structures within a simple and manageable framework providing a benchmark for rational behavior. The hierarchical structure embedded in the standard Investment Game naturally recalls the relationship between a Principal and an Agent in an agency framework. The game was played with integer numbers and this has allowed to keep the computational side of the task easy enough to be afforded even without the support of an electronic calculator.

2.1 The Intention Detection Game

To introduce the game it is useful to present a real-life situation which reflects the basic strategic tension of the experimental game. An employer (Principal) has to decide whether to invest on the training of an employee (Agent) or not. The decision is risky for the employer because there are no options to bound the employee's action after the training has been completed. The employee has thus the incentive to exploit the opportunity to be trained by the current employer and then be hired by another firm at a salary which accounts for the increase in productivity following the training but not for its cost. The current employer opens a vulnerability for herself by "sponsoring" the training of the employee. However, before entering this risky investment the employer is offered the opportunity of knowing the intentions of the employee, who is made aware of this option and of the choice of the Principal. The game, as depicted in Figure 1, is a very simple setting for strategic interaction between two Players, henceforth named Player I (Principal) and Player II (Agent). The game is a sequential game with complete and perfect information presented in an extensive form. It is also important to notice here that, *ex post*, there is no difference between a control on intentions or a control on actions as long as preferences are invariant with respect to the introduction of the monitoring system. However, from a more behavioral perspective and with the support of research in social psychology it is interesting to test for the effect of the introduction of the monitoring system.

Figure 1: The intention detection game



The first decisional node belongs to Player I who can decide whether to buy ($D = 1$) or not to buy ($D = 0$) a detection technology and pay accordingly to the choice undertaken a cost of c. In other words, the first mover can decide whether to enter a standard Investment Game or play a modified Investment Game with intention detection. To find the rational selfish equilibrium of the whole game, equilibrium in both subgames (IG with detection and without detection) will be separately considered.

When Player I chooses not to buy the technology both Players enter a standard Investment Game (IG). The IG is structured in the following way:

Player I, whom in the game is called Trustor, has an endowment E and decides how much to send to Player II, who is called Trustee. The condition $0 \leq a \leq E$ on the amount sent a must be fulfilled. The amount a is then multiplied by an exogenously given factor m and assigned to Player II. Player II and Player I are initially endowed with the same amount of wealth. The wealth of Player II after Player I's action amounts to $e + ma$. In the second stage of the game Player II chooses how much to submit to Player I. The amount sent, b, must satisfy the condition $0 \leq b \leq ma$. Finally, payoffs are computed and subjects are accordingly retributed.

Given the structure of the game, the value of the payoff for Player I, assuming for convenience linearity of the value function with respect to the payoffs, will thus be equal to

$$v_I = E - a + b$$

while the value for Player II will be equal to

$$v_2 = E + ma - b$$

Applying backward induction it is easy to show that the pure strategy subgame perfect equilibrium in the IG is for Player II to return an amount equal to 0 and thus for Player I to invest an amount equal to 0. Given these strategies, the resulting equilibrium outcomes in the investment subgame will be $v_I(s_I^*) = E$ and $v_{II}(s_{II}^*) = E$.

When Player I decides for detection the two actors interact over a structure which is the one of the Investment Game but the "meaning" of their actions is altered by the detection "technology". Indeed, when the detection strategy is purchased at cost c intentions on actions of Player II are perfectly visible to Player I and Player II is made aware of this. The word intention here means anticipated actions. Before being revealed by the detection strategy intentions are private information on the reciprocation nature of the partner. Player II knows that the detection technology is at work but does not know whatthe investment preferences of the Principal are. The detection of intentions is implemented in the following way: when the Principal buys the detection technology the Agent has to state for each possible action a of the Principal, where $(0 < a \leqslant (e - c))$, the intention to reward the investment made by the Principal. The collection of the complete strategy of the Agent is obtained through a strategy method: the output of the intention's scan is registered into a vector where each element of the vector corresponds to the reward in correspondence to each possible amount submitted by Player I. Given the parameters employed in the experiment the vector will have the following structure $\{b|1, b|2, b|3, b|4, b|5, b|6, b|7, b|8, b|9\}$, where the conditional term is the amount a submitted by Player I. After receiving information about intentions to reward the investment, Player I has to choose how much actually she wants to submit to Player II. The Principal's decision, a, must satisfy the condition $[0 \leqslant a \leqslant e - c]$. The actual investment decision is then matched with the corresponding element of the repayment vector. The payoffs of the Agent and the Principal are thus given by the following equations:

$$v_1 = E - c - a + b|a$$

$$v_2 = E + ma - b|a$$

To find the equilibrium outcome in the modified Investment Game with intention detection it is useful to start from the consideration that Player II, the Agent, can direct the strategy of Player I by appropriately chosing a strategy vector. The payoff structure implies that the gains of Player II increase in a linear way according to the factor m in Player's I investment while decrease by factor 1 with respect to the conditional payment, or effort, allocated in the game. As long as $m > 1$ the dominating action for Player II is to offer the minimum payment, or effort, that will suffice to induce Player I to invest her entire endowment at that stage ($a = E - c$). In order to induce the desired action the actual best solution in the game for Player II is thus to choose a conditional repayment vector equal to $\{E - c + \epsilon|9, b|8, ..., b|1\}$ where for each conditional $a \neq 9$ the upper bound condition $b < E - c + \epsilon$ must be respected. Given that the game is played with discrete values ϵ will be equal to 1 and thus, given the other parameters, $e - c + \epsilon$, where ϵ is an arbitrarily small value, will be equal to 10. Given the strategies above described, the couple of equilibrium outcomes is $v_I(s_I^*) = E$ and $v_{II}(s_{II}^*) = (E - c)m$

After having characterized equilibrium strategies in the two subgames originating from the detection decision, it emerges that both the identified equilibrium outcomes follow from subgame perfect strategies also when the whole game is considered. Player I will thus be indifferent, in equilibrium, between choosing the detection and the non detection strategy. The only discriminating element between the two alternative actions is the fact that the equilibrium outcome following from detection leads to a payoff allocation which is Pareto superior than the one following from non-detection[1].

To summarize, when Agents are rational and their preferences are oriented towards self-seekingness the game has two distinct equilibria in pure strategies and Player I will be indifferent, in equilibrium, between a detection and non-detection policy.

2.2 A simple model of reciprocity and intrinsic motivations

Recent contributions have focused on individual preferences which are not characterized only by self interest but take into account also the well-being of others. Aim of these works is to reconcile the standard economic approach of maximization of a given objective function with some puzzling findings of experimental works. On one side, there are models focused on distributional preferences (among others, Bolton and Ockenfels (2000); Fehr and Schmidt (1999)). On the other side, there are models that refer to Psychological Game Theory (Geanakoplos *et al.*, 1989) and are more focused on the intentions embedded in observed actions (among others, Rabin (1993); Falk and Fischbacher (2000)). Relying on the assumption that trust and reciprocity are the main determinants of behavior in a game of the

[1]given the experimental parameters the equilibrium outcomes in the non-detection are $v_I^* = 10; v_{II}^* = 10$ and those in the detection setting are $v_I^* = 10; v_{II}^* = 27$

kind considered here we will focus on a simple model of reciprocation in an Investment Game. The basic assumption of the value function presented below is that reciprocation has an autonomous psychological value[2]. The value function of Agents is accordingly enlarged to account also for this value component which is added to the monetary component of the value function.

$$V_2 = E - b + ma + \beta b - (\beta b - \alpha a)^2 \tag{1}$$

where

E = initial endowment

b = transfer to Player I

a = transfer from Player I

m = multiplier factor in the IG

β = measure of psychological value of reciprocity

α = measure of perceived kindness

The value component $e - b + ma$ represents the utility deriving from the monetary transfers in the game and is the same as in the rational selfish specification. The additional component $\beta b - (\beta b - \alpha a)^2$ captures the psychological value embedded in the game and links the utility from reciprocation to the perceived kindness of partner's action. The positive element βb represents the psychological gain of the subject when reciprocating action

[2]for a recent work dealing with the neurophysiological evidence of "psychological" value of reciprocation see Sanfey *et al.* (2003)

of Player I. The quadratic component $(\beta b - \alpha a)^2$ of the psychological value implies that both when monetary reciprocation is bigger than the perceived fairness or lower than this measure the subjects register a cost. The bigger the difference, the higher the psychological cost afforded is. A crucial element of the value function proposed is the parameter α which captures perceived kindness embedded in the action of Player I. Given the concavity of V_2 with respect to the amount returned b, the best reply function following from the first order condition $\frac{\partial V_2}{\partial b} = 0$ is

$$b^* = \frac{-1 + \beta + 2a\alpha\beta}{2\beta^2}$$

Turning to the rational decision of Player I, from the condition $\frac{\partial\left(E - a + \frac{-1+\beta+2a\alpha\beta}{2\beta^2}\right)}{\partial a} > 0$ one can obtain the condition upon which Player I will submit all of her endowment to Player II. A simplifying assumption which is introduced at this point is that psychological value from reciprocation equals monetary cost of reciprocating. Thus, under the assumption that $\beta = 1$ the best reply for Player II becomes $b^* = \alpha a$. This in turn implies that the condition for Player I to submit all her endowment to Player II is $\alpha > 1$. This can be interpreted in the following way: when the value of wealth transferred to Player II is perceived by Player II as higher than the monetary value of this amount the Player I has a "rational" and selfish incentive to invest all of her endowment even if there are no collateral available. A methodological assumption adopted in the specification of value functions is that the value function of Player I is a function exclusively of monetary payoffs. This assumption does not alter the nature of behavior described below and one can easily think of the possibility of introducing a psychological component which is increasing in the sum recturned by Player II. An additional value component of this kind simply increases the "utility" Player I obtains from an interaction based on trust and reciprocity but does not alter the game strategy.

When beliefs on α are correct, and in equilibrium we assume they are, one can provide the following characterization of the behavior of Player I as conditional upon α.

$$\begin{array}{ll} a^* = E & \text{if } \alpha > 1 \\ a^* = 0 & \text{if } \alpha < 1 \\ a^* \in [0, E] & \text{if } \alpha = 1 \end{array}$$

The above presentation of equilibrium behavior when Agents are characterized by value function 1 was based on an α conceived as an exogenously given factor. However, in order to experimentally test the impact of bounding strategies imposed by the Principal on the Agent, α will be endogenously determined by the action undertaken by the Principal in correspondence to the first decisional node. Specifically, $\alpha(D)$ will be lower than one when the Principal decides to purchase the detection strategy and bigger than one when the Principal adopts a non-detection strategy:

$$\begin{array}{ll} \alpha < 1 & \text{if } D = 1 \\ \alpha > 1 & \text{if } D = 0 \end{array}$$

The way α is modeled captures the intrinsic motivations component of the game. Relying on previous contributions it has been shown in section

1 that in environments where autonomy in decision making is preserved, actions are more likely to be characterized by intrinsic motivations than in environments where actions are bounded by the Principal. It has also been shown that Agents characterized by intrinsic motivations are more likely to foster cooperative behavior. These two observation give support to the functional form adopted for α. An $\alpha > 1$ means that the mount of money submitted by the Principal has an higher value to the Agent than its mere monetary value because it comprises also a kindness element related to the decision to warrant autonomy to the Agent. An $\alpha < 1$ means that Agents attach a negative component to the monetary transfer from Player I deriving from the control practices implemented.

Given equation 1 and the functional from of α, the Principal's equilibrium actions in the game will be to refuse to buy the detection technology ($D = 0$) and to invest her entire wealth in the Investment Game ($a = E$). On the Agent's side, the action undertaken in equilibrium will be to return an amount equal to $E \times \alpha$. Thus, equilibrium outcomes are, for Player I and Player II respectively, $V_1* = \alpha E$ and $V_2* = E + ma$. It is interesting to notice that this outcome is Pareto preferable with respect to the two equilibria obtained under the selfish rational assumption.

The aalue functions specified above allow to explore two alternative research hypotheses within our simple experimental game. The first hypothesis, which can be termed *rational selfishness*, is that Principals will either choose to detect the action of the Agents who will reply with minimum effort or that Principals will not detect Agent's intentions but submit an amount equal to zero in the Investment Game. The second hypothesis, which can be termed *intrinsically motivated rational reciprocity*, is that Principals will reject the opportunity to detect intentions of the Agent and "invest" their entire endowment in the Investment Game. Moreover, Agents are expected to return an amount of wealth proportional to α which however is higher than the amount received from Player I. The game allows to clearly identify whether these hypotheses are supported by the behavior in the game and thus to draw some conclusions about nature of observed behavior.

To conclude this section it is worthwhile to point out an implicit assumption of the value specification 1 described above is that the game considered frames attention of the Players on reciprocity issues while distributional considerations are neglected. This assumption is related to the dynamic nature of the game that naturally tends to focus more on transactions than on final distribution[3]. Future research may try to expand the model to encompass also distributional concerns.

2.3 Experimental Procedures

The computer-based experiment was run at the Computational and Experimental Economics Laboratory (CEEL) of the University of Trento[4] on a

[3]for an insight into the interplay between distributional and reciprocal concerns in the Investment Game one can refer to Cox (2004)

[4]financial support was provided by CEEL which is acknowledged for this but also for the technical support provided before, during and after the experiment. Special thanks are due to Professor Luigi Mittone, Ivan Soraperra and Marco Tecilla but also to the staff of the laboratory which provided a valuable contribution to this paper. The value of the support of Dominique Cappelletti is incommensurable and not limited to the present paper. The financial and technical support of CEEL is gratefully acknowledge

client-server infrastructure purposely built using the programming language Borland© Delphi©. Participants were undergraduate students of the University of Trento. The majority of them were students of Economics. Two identical sessions were run on December 16^{th}, 2004, with 20 participants per session, for a total of 2x20=40 Subjects. Instructions were read aloud before the game started and participants were free to ask for clarifications after having read the instructions. Each participants was endowed, independently of the role in the experiment, with 10 units of Experimental Currency Unit (ECU). The exchange rate between ECU and Euro reported on the instructions sheet was $0,75$ € for each ECU. The payment was made available in cash immediately after the second, and last, session. Anonymity among participants was warranted during and after the game and matching between the subjects was randomly determined. A control questionnaire was implemented before the start of the game in order to prevent noise in the outcome due to a misunderstanding of the experimental procedures. The questionnaire was implemented through an electronic form which had to be filled with correct answers on different aspects of the game. It is important to remark here that the experiment was run under a condition of zero socialization. Matching was random and anonymous, subjects had no chance to interact after having entered the lab. The experiment took on average 40 minuts and the average payment was about 10 €. A brief questionnaire on some aspects of the experiment was handled out to the subjects at the end of the experiment in order to better understand observed choices.

3 Data Analysis

The first pattern in the data that has to be considered is the number of subjects who decide to buy or not buy the detection technology. 85% of the participants (17 out of 20) chose the detection policy. This simple descriptive statistics already rules out the possibility that Player I believes in positive reciprocity as a response to the decision of do not detect Agent's action. Moreover, it suggests that among the two rational equilibrium strategies the Player prefer to choose the one prescribing the purchase of the detection technology.

For what attains outcomes in the non detection framework it must be considered that the small number of Agents observed (i.e., 3) strongly limits the analysis of behavior in the Investment Game.

Table 1: Non detection framework

	N	**mean**	**med**	**sd**	**min**	**max**
Trustor (Player I)	3.00	7.00	7.00	3.00	4.00	10.00
Trustee (Player II)	3.00	0.11	0.14	0.10	0.00	0.20

Table 1 illustrates the decision of the trustor and the trustee in the standard Investment Game. The decision of the Trustee is reported as a percentage of the amount sent by the Trustor. What emerges from the table is that the average contribution of the Trustor is 70% of the total endowment available with the lower contribution equal to 4 and the higher equal to the total endowment. The magnitude of trust is on average higher than what observed in previous works (e.g., Berg *et al.*, 1995) and this seems to suggest that those who do not choose to detect Agent's action are "optimistic" about the reciprocatory attitude of their partner. Actual choices to reciprocate are, contrary to what expected, much lower than values registered in the literature on Investment Game. Agents do not react positively to the decision to trust undertaken by the Principal and try to almost fully exploit the vulnerability the Principal opens for herself.

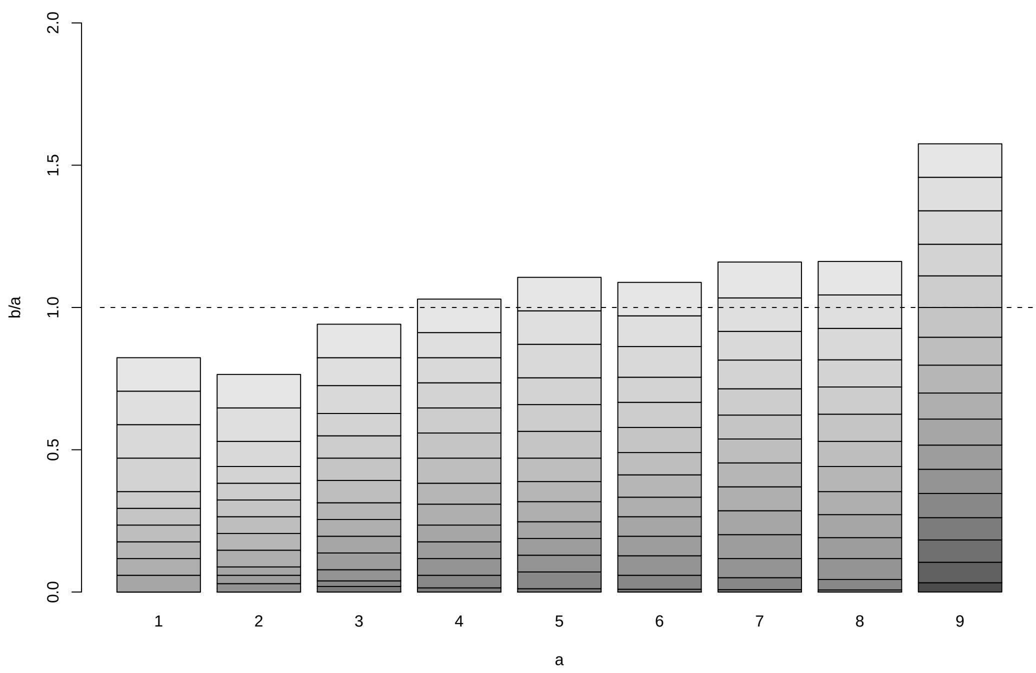
When considering the average monetary gains of Player I and Player II reported in Table 2, it emerges that Players II gain considerably more than Players I thanks to their opportunistic behavior.

Table 2: Payoffs in the non-detection environment

	N	**mean**	**med**	**sd**	**min**	**max**
Trustor (Player I)	3.00	4.00	4.00	2.00	2.00	6.00
Trustee (Player II)	3.00	30.00	30.00	8.00	22.00	38.00

In general what emerges from observations referring to autonomy supportive environments is that the self-imposed absence of bounds on Agent's actions undertaken by the Principal is not rewarded by the Agents who not only do not show reciprocity but apparently neglect also any consideration in terms of fairness. Moving the attention to the autonomy bounding configuration, Figure 2 reports the pattern of repayment for Player II for each possible amount sent by Player I in case of detection. The possible offers of Player I are reported on the $X - axis$ while on the $Y - axis$ the top of the bar measures the average level of the capital after the investment in terms of percentage on the capital invested (i.e., b/a). Each cell of the bar represents the individual share of the mean as measured by the top of that bar. It is important to recall here that in a rational equilibrium perspective choices different from $a = 9$ do not meet the incentive compatibility requirement. Graph 2 provides an evidence of the intention to repay the investment for each amount between 1 and 9 that Player I can decide to submit.

Figure 2: Detection vector



The dashed line in the graph corresponds to the threshold that separates positive investments from negative investments. For investments corresponding to $a < 4$ the average gain is negative while it becomes positive for investments greater than 3. Observing the sharing of the average, it emerges that as the possible amount invested increases the number of cells increases. This has to be interpreted as due to the fact that a loer amount of Players II return 0 to Players I. The general tendency evidenced by the data is a monotonic increase in correspondence to increase in potential investment. From graph 2 it emerges a steep increase in repayment in correspondence to potential investment equal to 9. Differentiated behavior in correspondence to this measure with respect to other measures of potential investment is confirmed also by the non parametric two sample Wilcoxon tests reported in Table 3.

Table 3: Wilcoxon test (p-values)

a	1	2	3	4	5	6	7	8
9	0.008	0.000	0.001	0.007	0.036	0.012	0.059	0.104

Differences are statistically significant at the 10% condifence level in all the cases except than in the comparison between potential investment equal to 9 and 8 (i.e., last cell of Table 3).

From Figure 2 and the tests performed one can infer that no punishing behavior has been shown by Player II as a response to the decision of Player I to introduce the detection system. Despite the introduction of the autonomy bounding strategy the repayment in correspondence to the equilibrium choice of Player I is considerably higher that the repayment in the other conditional choices. From this observation it is possible to conclude that no retaliation is implemented by Player II after the introduction of the detection. For what attains the behavior of Player I in the detection situation, this is in general in line with equilibrium strategy which prescribes to send all the endowment. For what attains out of equilibrium behavior one observation to submit 0, one observation to submit three and, finally, two observations to submit 8 have been observed. The meaning of out of equilibrium choices of Player II is difficult to be interpreted, in particular for what attains low choices. Mistakes are the best candidates to explain these deviations.

A perspective on the dynamic of the game could be gathered also focusing on the final payoffs of the two Players which are reported in Table 4

Table 4: Payoffs in the detection environment

	N	**mean**	**med**	**sd**	**min**	**max**
Player I	17.00	14.35	14.00	2.76	9.00	18.00
Player II	17.00	20.65	22.00	3.72	10.00	25.00

As it emerges from Table 4, Player II, on average ends the game with a considerably higher endowment than Player I. This is in line with the equilibrium outcome but the registered values are different than those expected in equilibrium. In particular, Player I gains more in the actual game than in the forecasted equilibrium allocation. As the two Players share a given efficiency gain the opposite is necessarily true for Player II.

It is of some interest to consider the structure in the data of the strategy vector of Player II when the detection technology is at work. As remarked above, the equilibrium choice is to send one unit more than the amount sent by Player I in correspondence to the highest possible amount that Player I can send. Given this, all the other values of the vector are "meaningless". Different from what expected, however, a well defined pattern in the data referring to out of equilibrium strategy is observed. The structure in the data suggests that values are not chosen randomly and data collected will be considered as a survey on reciprocity propension in a detection condition.

Figure 3: Player II: out of equilibrium choices in the intention vector

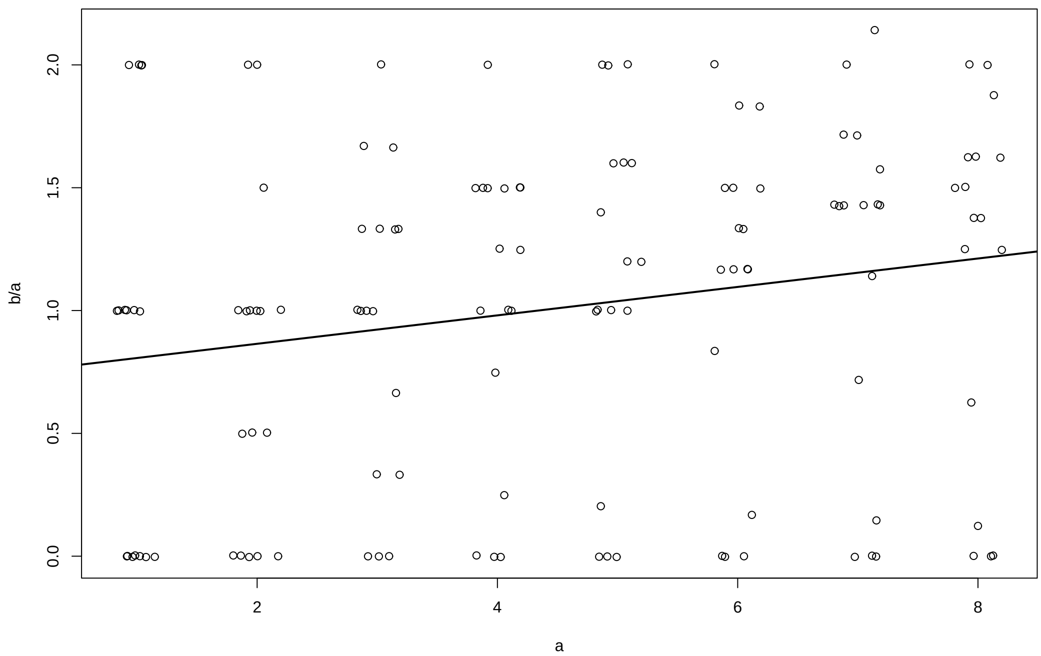


Figure 3 depicts the retributed capital for each out-of-equilibrium potential investment ($b/a \sim a$). A small noise is added to values to render the representation on the plot more effective. The continuous line represents a linear regression fitted with ordinary least squares. The inclination of the line is positive and equal to 0.06 and is statistically significant at the conventional 5% level but the variation explained by the model is very low ($R^2 = 0.03$). What emerges from the regression is that the higher the amount sent the more Subjects tend to "hypothetically" reward wealth received. This pattern does not emerge among the few observations collected in the "real" Investment Game and this suggests that retribution is meaningful in trust decision and non-rewarded decision of reciprocation may be upward biased.

4 Conclusions

The hypotheses about reciprocity and autonomous-supportive behavior are strongly rejected because the observed behavior of the Players is very far from the predicted behavior under this assumption. Only 3 subjects out of 20 chose not to detect the activity of Agents and sent a relevant amount of their wealth to the partner but only one of them sendt all the endowment.

For what attains the rational selfish hypothesis it must firstly be noticed that final allocations of payoffs observed are, on average, not far from what prescribed by the hypothesis. An interesting deviation from the equilibrium strategy has been registered in the behavior of Player II. Rational self-regarding preferences will prescribe a minimal repayment in correspondence of the maximum amount sent by Player I. Instead, that has been observed, with statistical significance, is that the repayment choices do not differ considerably in the conditional decisions different from $c|9$ and, at the same time, decision in $c|9$ are statistically different from the large part of other conditional decisions. The difference derives from a mark-up on the repayment which cannot be explained in the standard perspective of rational self regarding Agents. Looking at the intention vector it emerges that even if the decision out of equilibrium are "meaningless" subjects tend to follow a linear reciprocity repayment with a repayment ratio slightly above 1 on average. In order to better understand the intentions behind the registered mark-up it could be useful, even if conditions are very different, to refer to the evidence emerging from the Ultimatum Game (Güth *et al.*, 1982). No evidence of costly-punishment has been registered in the decisions of Player II when the detection condition was overimposed to her choices. Considering the evidence of altruistic punishment in previous contributions based on games involving social norms (e.g., Fehr and Gachter, 2002; Fehr and Fischbacher, 2004) this suggests that in our setting the introduction of a monitoring technology has not been perceived as a constraint or as a threat to cooperation by the monitored Agents. At a deeper level of analysis the absence of costly punishment might be attributed to two distinct factors. The subjects may not be endowed with reciprocity concerns or, alternatively, the introduction of a detection strategy has no impact on reciprocity concerns. Further research is needed to disentangle these two different sources of non reciprocity in the game.

For what attains the few observations associated with the non detection setting, it must be observed that they do not conform to any of the research hypothesis reported above. Trustors invest an amount which is on average higher than what usually observed in Investment Games. Player II's behavior is almost fully opportunistic and leads to big losses for Players of type 1. This suggests that the behavior of Player II is at all influenced by the trust signaling monitoring decision of Player I. All the investments undertaken by Players 1 who do not buy the detection strategy register considerable losses. For what attains the final allocation of payoffs a big gap between Player I and Player II emerges in the detection environment while the situation is more balanced in the non-detection setting, however also in the latter Player II are made relatively better off in comparison with Players of type 1. The fact that all the Principals acting according to an equilibrium strategy choose the detection technology might signal that they are concerned with collective welfare and are not characterized by preferences for equity in the payoff distribution. It must however be noticed that when Player's type is not common knowledge it is safer for Player II to choose the equilibrium which prescribes of not buying the detection technology and send an amount equal to zero in the Investment Game.

An alternative hypothesis that deserves more attention in future research is that Principals are attracted by the monitoring technology and focus more on the potential power of control and do not asess correctly the consequences of their actions.

The outcomes of the experiment sharply differ from previous field studies, mainly based on self-reported values, which have been briefly considered in the introduction. The difference in identification with an organizational structure between experiments and real life situations represents a candidate explanation for different pattern observed (Simon, 1991). A direction for future research will be to consider the impact of "socialization" on the game introducing a 2x2 design, where the two factors will be ex-ante non-strategic interaction (cheap talk, common task ...) and repetition of the game.

To briefly summarize, results of the experiment it must be evidenced how Principals opting for the detection strategy conclude the game with a positive gain while Principals opting for a more autonomous-supportive framework end the game with losses on their initial endowment. Moreover, not only Principals do not seem to believe in the potential reciprocator attitude of the Agents but the few who do are "betrayed" by Agents revealing to be self-seeking oriented and who not show any kind of reciprocity. Moreover, Agents acting under a detection regime provide a return on the investment of the Principal that is higher than that prescribed by rational equilibrium strategy. All the evidence collected is in favor of control practices in interactions similar to those captured by the game. Principals do not pursue autonomy in decision making or the Agents and in th efew cases when this happens no positive impact on Agents' reciprocity is registered.

References

Arrow, K. J. (1974). *The Limits of Organizations*. Norton Company.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior*.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity and competition. *The American Economic Review*, **90**(1).

Buchan, N. and Croson, R. (2004). The boundaries of trust: own and others actions in the us and china. *Journal of Economic Behavior and Organization*, **55**.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*.

Deci, E. L., Connell, J. P., and Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, **74**(4), 580–590.

Falk, A. and Fischbacher, U. (2000). A theory of reciprocity. Institute for Empirical Research in Economics University of Zurich-Working Paper No. 6.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, (25), 63–87.

Fehr, E. and Gachter, S. (2002). Altruistic punishment in humans. *Nature*.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *The Quarterly Journal of Economics*.

Fehr, E., Kirchler, E., Weichbold, A., and Gächter, S. (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics*, **16**(2), 324–351.

Gächter, S., Herrmann, B., and Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior and Organization*, **55**, 505–531.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, pages 60–79.

Glaeser, E., Laibson, D. I., Scheinman, J. A., and Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic Behavior and Organization*, **3**, 367–388.

Güth, W., Levati, M. V., and Ploner, M. (2004). The effect of group identity in an investment game. MPIEW Discussion Paper.

Hardin, R. (2001). *Conceptions and Explanations of Trust*, chapter 1, pages 3–37. Russell Sage Foundation.

Kramer, R. M. (1999). *Annual Review of Psychology*.

March, J. G. (1994). *A Primer on Decision Making*, chapter 2. The Free Press.

Messik, D. M. and Kramer, R. (2003). Trust as a form of shallow morality. In K. S. Cook, editor, *Trust in Society*, chapter 3, pages 89–117. Russell Sage Foundation.

Porta, R. L., de Silanes, F. L., Shleifer, A., and Vishny, R. W. (1997). Trust in large organizations. *The American Economic Review*, **87**(2), 333–338.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, pages 7–63.

Rabin, M. (1993). Incorporting fairness into game theory and economics. *The American Economic Review*, **83**(5), 1281–1302.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, **25**, 54–67.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, **300**.

Simon, H. A. (1991). Organizations and markets. *The Journal of Economic Perspectives*, **5**, 25–44.

Williamson, O. E. (1993). Calculativenesss, trust and economic organization. *Journal of Law and Economics*, **XXXVI**.

Yamaghishi, T. (2001). *Trust in Society*, chapter 4, pages 121–147. Russell Sage Foundation.